Five endometrial cancer risk loci identified through genome-wide association analysis

Five endometrial cancer risk loci identified through genome-wide association analysis

Timothy HT Cheng[1]*, Deborah J Thompson[2]*, Tracy A O'Mara[3], Jodie N Painter[3], Dylan M Glubb[3], Susanne Flach[1], Anabelle Lewis[1], Juliet D French[3], Luke Freeman-Mills[1], David Church[1], Maggie Gorman[1], Lynn Martin[1], National Study of Endometrial Cancer Genetics Group (NSECG)[1], Shirley Hodgson[4], Penelope M Webb[3], The Australian National Endometrial Cancer Study Group (ANECS)[3], John Attia[5, 6], Elizabeth G Holliday[5, 6], Mark McEvoy[6], Rodney J Scott[5, 7-9], Anjali K Henders[3], Nicholas G Martin[3], Grant W Montgomery[3], Dale R Nyholt[3, 10], Shahana Ahmed[11], Catherine S Healey[11], Mitul Shah[11], Joe Dennis[2], Peter A Fasching[12, 13], Matthias W Beckmann[13], Alexander Hein[13], Arif B Ekici[14], Per Hall[15], Kamila Czene[15], Hatef Darabi[15], Jingmei Li[15], Thilo Dörk[16], Matthias Dürst[17], Peter Hillemanns[18], Ingo Runnebaum[17], Frederic Amant[19], Stefanie Schrauwen[19], Hui Zhao[20, 21], Diether Lambrechts[20, 21], Jeroen Depreeuw[19-21], Sean C Dowdy[22], Ellen L Goode[23], Brooke L Fridley[24], Stacey J Winham[23], Tormund S Njølstad[25, 26], Helga B Salvesen[25, 26], Jone Trovik[25, 26], Henrica MJ Werner[25, 26], Katie Ashton[5, 8, 9], Geoffrey Otton[27], Tony Proietto[27], Tao Liu[28], Miriam Mints[29], Emma Tham[28], RENDOCAS[28], CHIBCHA Consortium[1], Mulin Jun Li[30], Shun Yip[30], Junwen Wang[30], Manjeet K Bolla[2], Kyriaki Michailidou[2], Qin Wang[2], Jonathan P Tyrer[11], Malcolm Dunlop[31, 32], Richard Houlston[33], Claire Palles[1], John L Hopper[34], AOCS Group[3, 35], Julian Peto[36], Anthony J Swerdlow[33, 37], Barbara Burwinkel[38, 39], Hermann Brenner[40-42], Alfons Meindl[43], Hiltrud Brauch[42, 44, 45], Annika Lindblom[28], Jenny Chang-Claude[46], Fergus J Couch[23, 47], Graham G Giles[34, 48, 49], Vessela N Kristensen[50-52], Angela Cox[53], Julie M Cunningham[23, 47], Paul D P Pharoah[11], Alison M Dunning[11], Stacey L Edwards[3], Douglas F Easton[2, 11+], Ian Tomlinson[1+], Amanda B Spurdle[3+]

* contributed equally to this work
+ to whom correspondence should be addressed

[1] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.
[2] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK.
[3] Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, 4006, Australia.
[4] Department of Clinical Genetics, St George's, University of London, London, SW17 0RE, UK.
[5] Hunter Medical Research Institute, John Hunter Hospital, Newcastle, NSW, 2305, Australia.
[6] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, NSW, 2305, Australia.
[7] Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, 2305, Australia.
[8] Centre for Information Based Medicine, University of Newcastle, NSW, 2308, Australia.
[9] School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, NSW, 2308, Australia.
[10] Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, 4006, Australia.
[11] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK.
[12] University of California at Los Angeles, Department of Medicine, Division of Hematology/Oncology, David Geffen School of Medicine, Los Angeles, CA, 90095, USA.
[13] Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, 91054, Germany.
[14] Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, 91054, Germany.

[15] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE-171 77, Sweden.
[16] Hannover Medical School, Gynaecology Research Unit, Hannover, 30625, Germany.
[17] Department of Gynaecology, Jena University Hospital - Friedrich Schiller University, Jena, 07743, Germany.
[18] Hannover Medical School, Clinics of Gynaecology and Obstetrics, Hannover, 30625, Germany.
[19] Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, University Hospitals, KU Leuven - University of Leuven, 3000, Belgium.
[20] Vesalius Research Center, Leuven, 3000, Belgium.
[21] Laboratory for Translational Genetics, Department of Oncology, University Hospitals Leuven, Leuven, 3000, Belgium.
[22] Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, Mayo Clinic, Rochester, MN, 55905, USA.
[23] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905, USA.
[24] Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, 66160, USA.
[25] Centre for Cancerbiomarkers, Department of Clinical Science, The University of Bergen, 5020, Norway.
[26] Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, 5021, Norway.
[27] School of Medicine and Public Health, University of Newcastle, Newcastle, NSW, 2308, Australia.
[28] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, SE-171 77, Sweden.
[29] Department of Women's and Children's Health, Karolinska Institutet, Karolinska University Hospital, Stockholm, SE-171 77, Sweden.
[30] Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.
[31] Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.
[32] MRC Human Genetics Unit, Western General Hospital Edinburgh, Edinburgh, EH4 2XU, UK.
[33] Division of Genetics and Epidemiology, Institute of Cancer Research, London, SM2 5NG, UK.
[34] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Vic, 3010, Australia.
[35] Peter MacCallum Cancer Center, The University of Melbourne, Melbourne, 3002, Australia.
[36] London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK.
[37] Division of Breast Cancer Research, Institute of Cancer Research, London, SM2 5NG, UK.
[38] Molecular Biology of Breast Cancer, Department of Gynecology and Obstetrics, University of Heidelberg, Heidelberg, 69120, Germany.
[39] Molecular Epidemiology Group, German Cancer Research Center, DKFZ, Heidelberg, 69120, Germany.
[40] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany.
[41] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, 69120, Germany
[42] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany
[43] Department of Obstetrics and Gynecology, Division of Tumor Genetics, Technical University of Munich, Munich, 80333, Germany.
[44] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, 70376, Germany.
[45] University of Tübingen, Tübingen, 72074, Germany
[46] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany.
[47] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, 55905, USA.

[48] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Vic, 3004, Australia.
[49] Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Vic, 3004, Australia.
[50] Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, 0310, Norway.
[51] The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, 0316, Norway .
[52] Department of Clinical Molecular Oncology, Division of Medicine, Akershus University Hospital, Lørenskog, 1478, Norway.
[53] Sheffield Cancer Research, Department of Oncology, University of Sheffield, Sheffield, S10 2RX, UK.

Abbreviations: EC, endometrial cancer; CI, confidence interval; GWAS, genome-wide association study; LD, linkage disequilibrium; OR, odds ratio; kb, kilobase; Mb, megabase; PCA, principal components analysis; DHS, DNase1 hypersensitivity site.

Abstract

We conducted a meta-analysis of three endometrial cancer (EC) GWAS and two replication phases totaling 7,737 EC cases and 37,144 controls of European ancestry. Genome-wide imputation and meta-analysis identified five novel risk loci at genome-wide significance at likely regulatory regions on chromosomes 13q22.1 (rs11841589, near *KLF5*), 6q22.31 (rs13328298, in *LOC643623* and near *HEY2* and *NCOA7*), 8q24.21 (rs4733613, telomeric to *MYC*), 15q15.1 (rs937213, in *EIF2AK4,* near *BMF*) and 14q32.33 (rs2498796, in *AKT1* near *SIVA1*). A second independent 8q24.21 signal (rs17232730) was found. Functional studies of the intergenic 13q22.1 locus showed that rs9600103 (pairwise r^2=0.98 with rs11841589) is located in a region of active chromatin that interacts with the *KLF5* promoter region. The rs9600103-T EC protective allele suppressed gene expression *in vitro* suggesting that the regulation of *KLF5* expression*,* a gene linked to uterine development, is implicated in tumorigenesis. These findings provide enhanced insight into the genetic and biological basis of EC.

Endometrial cancer (EC) is the fourth most common cancer in women in the United States[1] and Europe[2], and the most common cancer of the female reproductive system. The familial relative risk is ~2[3,4], but highly penetrant germline mutations in mismatch repair genes[5], and DNA polymerases[6,7] account for only a small proportion of the familial aggregation. Our previous GWAS and subsequent fine-mapping identified the only two reported genome-wide significant EC risk loci, tagged by rs11263763 in *HNF1B* intron 1[8] and rs727479 in *CYP19A1* intron 4[9].

To identify additional EC risk loci, we re-analysed data from our previous GWAS (ANECS, SEARCH datasets[10]) and conducted a meta-analysis with two further studies (**Supplementary Figure 1**). The first was an independent GWAS; the National Study of Endometrial Cancer (NSECG), including 925 EC cases genotyped using the Illumina 660W array, 1,286 cancer-free controls from the CORGI/SP1 GWAS[11,12] and 2,674 controls from the 1958 Birth Cohort[13]. The second study comprised 4,330 EC cases and 26,849 controls from Europe, the United States and Australia, genotyped using a custom array designed by the Collaborative Oncological Gene-environment Study (COGS) initiative[14–17] (**Supplementary Table 1, Supplementary Note**).

We first performed genome-wide imputation using 1000 Genomes Project data, allowing us to assess up to 8.6 million variants with allele frequency ≥1% across the different studies. Per-allele odds ratios and P-values for all SNPs in the GWAS and iCOGS were obtained using a logistic regression model. There was little evidence of systematic overdispersion of the test statistic (λ_{GC}=1.002-1.038, **Supplementary Figure 3**). A fixed-effects meta-analysis was conducted for all 2.3 million typed and well-imputed SNPs (info score>0.90) in a total of 6,542 EC cases and 36,393 controls. The strongest associations were with SNPs in LD with previously identified EC risk SNPs in *HNF1B*[10,8,18] and *CYP19A1*[9,19] (**Figure 1, Table 1**). For fourteen 1.5Mb regions containing at least one novel SNP with $P_{meta}<10^{-5}$, we performed regional imputation using an additional reference panel that comprised 196 high-coverage whole genome-sequenced UK individuals (**Supplementary Table 2**).

Five novel regions containing at least one EC risk SNP with $P_{meta}<10^{-7}$ were identified and the most strongly associated SNP in each region was genotyped in an additional 1,195 NSECG EC cases and 751 controls using competitive allele-specific PCR (KASPar, KBiosciences) and the Fluidigm BioMark System (**Supplementary Table 3**). Duplicate samples displayed concordance >98.5% between different genotyping platforms (**Supplementary Table 4**). All five SNPs were associated with EC at genome-wide significance ($P<5\times10^{-8}$, **Table 1, Figure 2**), and these associations remained highly significant when analysis was restricted to cases

with endometrioid subtype only. Endometrioid-only analysis did not reveal any additional risk loci. eQTL analysis (**Online Methods**) in normal uterine tissue,[20] and EC tumour and adjacent normal tissue[21] did not yield any SNPs robustly associated with the expression of nearby genes at the EC risk loci (**Supplementary Table 7)**. However, for each risk locus, bioinformatic analysis including cell-type-specific expression and histone modification data identified correlated SNPs within 500kb in likely enhancers and multiple potential regulatory targets (**Supplementary Table 6, Supplementary Figure 5**). The most compelling candidates for future functional analysis are described below.

rs13328298 (OR=1.13, 95%CI:1.09–1.18, $P=3.73\times10^{-10}$) on 6q22.31 lies in the long non-coding RNA *LOC643623*, 54kb upstream of *HEY2* and 86kb upstream of *NCOA7*. *HEY2* is a helix-loop-helix transcriptional repressor in the Notch pathway, which maintains stem cells, and dysregulation has been associated with different cancers[22]. *NCOA7* modulates the activity of the estrogen receptor via direct binding[23].

The second locus (rs4733613, OR=0.84, 95%CI:0.80–0.89, $P=3.09\times10^{-9}$) is at 8q24.21. Stepwise conditional logistic regression identified another independent signal in this region, rs17232730 (pairwise r^2=0.02, $P_{cond}=1.29\times10^{-5}$, **Table 2**). Both EC SNPs lie further from *MYC* (784-846kb telomeric) than most of the other cancer SNPs in the region, including those for cancers of the bladder[24,25], breast[26,16], colorectum[11,27], ovary[28] and prostate[29,30]. rs17232730 is in moderate LD with the ovarian cancer SNP rs10088218 (r^2=0.43), with both cancers sharing the same risk allele, but rs4733613 is not in LD ($r^2\leq0.02$) with any other cancer SNP in the region (**Supplementary Figure 5**). A role in tumorigenesis is implicated for several miRNAs in the region[31]. Of these, miR-1207-5p is reported to repress *TERT*, a locus also implicated in EC risk[32].

The lead SNP at 15q15 (rs937213; OR=0.90, 95%CI:0.86–0.93, $P=1.77\times10^{-8}$) lies within an intron of *EIF2AK4. EIF2AK4* encodes a kinase that phosphorylates EIF2α and downregulates protein synthesis during cellular stress[33]. Another nearby gene, *BMF,* encodes an apoptotic regulator moderately to highly expressed in glandular endometrial tissue[34].

At 14q42, the lead SNP rs2498796 (OR=0.89, 95%CI:0.85–0.93, $P=3.55\times10^{-8}$) lies in intron 3 of oncogene *AKT1,* which is highly expressed in the endometrium[34]. Several SNPs in LD with rs2498796 are bioinformatically linked with regulation of *AKT1* and four other nearby genes (*SIVA1, ZBTB42, ADSSL1* and *INF2*; **Supplementary Table 6, Supplementary Figure 5**). *AKT1* acts in the PI3K/AKT/MTOR intracellular signaling pathway, which affects cell survival and proliferation[35] and is activated in endometrial tumors[36], especially aggressive

disease[37,38,39]. *SIVA1* encodes an apoptosis regulatory protein that inhibits p53 activity[40,41] and enhances epithelial–mesenchymal transition to promote motility and invasiveness of epithelial cells[42]. *INF2* expression is reported to act as a promigratory signal in gastric cancer cells treated with mycophenolic acid[43].

The final novel EC SNP was rs11841589 (OR=1.15, 95%CI:1.11–1.21, $P=4.83\times10^{-11}$) on chromosome 13q22.1, 163kb and 445kb downstream from Kruppel-like factors *KLF5* and *KLF12,* respectively. KLF5 is a transcription factor associated with cell cycle regulation, and it plays a role in uterine development, homoeostasis and tumorigenesis[44–47]. Elevated *KLF5* levels are strongly correlated with activating *KRAS* mutations[48] and KLF5 is targeted for degradation by the tumor suppressor FBXW7. Both *FBXW7* and *KRAS* are commonly mutated in EC[49]. rs11841589 was one of a group of five highly correlated SNPs ($r^2 \geq 0.98$) surpassing genome wide significance in a 3kb LD block bounded by rs9600103 ($P=8.70\times10^{-11}$) and rs11841589 (**Figure 4a)**. There was no residual association signal at this locus ($P_{cond} > 0.05$) after conditioning for rs11841589. Bioinformatic analysis suggested that the causal variant at the intergenic 13q22.1 locus may affect a regulatory element that modifies *KLF5* expression (**Supplementary Figure 5**); rs9600103 overlaps a vertebrate conservation peak, and a DNaseI hypersensitivity site (DHS) in estrogen and tamoxifen-treated ENCODE[50] Ishikawa cells (**Figure 4a**). In addition, in a Hi-C chromatin capture experiment in Hela S3 cells[51], an interaction loop was observed between a segment containing the *KLF5* promoter and the rs11841589/rs9600103 locus (P=0.004, **Supplementary Figure 6**).

We further investigated the epigenetic landscape of a 16kb region around rs11841589 and rs9600103 that contained the SNPs most strongly associated with EC, by analysis of three EC cell lines: Ishikawa is homozygous for the rs9600103-A and rs11841589-G high-risk alleles, and provided a comparison with the ENCODE data; ARK-2 is homozygous for the low-risk T alleles at both SNPs; and AN3CA is a non-*KLF5* expressing line that is homozygous for the high-risk alleles (**Supplementary Figure 7)**. We conducted formaldehyde-assisted identification of regulatory elements (FAIRE, to identify regions of open chromatin), and chromatin immunoprecipitation (ChIP) using antibodies against H3K4Me2 (marker of transcription factor binding[52]) and panH4Ac (marker of active chromatin). Although the anti-H4Ac ChIP did not display a consistent signal in the region, signals from FAIRE and anti-H3K4Me2 ChIP were specifically present in the *KLF5*-expressing lines and were co-located with the conservation peak and DHS from the ENCODE data at rs9600103, providing strong evidence for open chromatin and transcription factor binding here (**Figure 4a**). We then conducted chromatin conformation capture experiments for the *KLF5*-expressing Ishikawa endometrial cancer cells and we found a significant interaction between the *Nco*I restriction

fragment containing the rs11841589/rs9600103 risk loci SNPs and the promoter region of *KLF5* (**Figure 4b**).

The regulatory nature of the region around rs9600103 and rs11841589 was investigated using allele-specific luciferase enhancer reporter assays in Ishikawa cells (**Figure 4c)**. Paired t-tests were used to compare the relationships between fragments containing the rs11841589 and rs9600103 alleles, and the pGL3-Promoter reporter vector (no insert) control (**Supplementary Table 8**). Fragments containing the rs9600103-T, rs11841589-T and rs11841589-G alleles had activity significantly lower than that of the pGL3-Promoter control (P≤0.014). In contrast, the construct containing the rs9600103-A risk allele had luciferase expression similar to the pGL3-Promoter control (P=0.23) and significantly higher than that of rs9600103-T (P=0.02), rs11841589-T (P=0.05) and rs11841589-G (P=0.04). These results suggest that the EC risk tagged by rs11841589 is at least partly due to a regulatory element containing rs9600103, which interacts with the *KLF5* promoter region, and the risk rs9600103-A allele is likely associated with increased gene expression.

In summary, this meta-analysis identified five novel EC risk loci at genome-wide significance, bringing the total number of common EC risk loci identified by GWAS to seven (**Figure 1)**. Together with other risk SNPs reaching study-wide significance[32,53,54], these explain ~1.6% of the EC familial relative risk. Novel EC risk SNPs lie in likely enhancers predicted to regulate genes or miRNAs with known or suspected roles in tumorigenesis, and we specifically showed that a functional SNP at 13q22.1 may sit within a transcriptional repressor of *KLF5*. Our findings further clarify the genetic etiology of EC, provide regions for functional follow-up, and add key information for future risk stratification models.

Method**s**

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Acknowledgments

The authors thank the many individuals who participated in this study and the numerous institutions and their staff who supported recruitment, detailed in full in the Supplementary Text.

The iCOGS endometrial cancer analysis was supported by NHMRC project grant [ID#1031333] to ABS, DFE and AMD. ABS, PW, GWM, and DRN are supported by the NHMRC Fellowship scheme. AMD was supported by the Joseph Mitchell Trust. IT is supported by Cancer Research UK and the Oxford Comprehensive Biomedical Research Centre. THTC is supported by the Rhodes Trust and the Nuffield Department of Medicine. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement no 223175 [HEALTH-F2-2009-223175] [COGS], Cancer Research UK [C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565], the National Institutes of Health [CA128978] and Post-Cancer GWAS initiative [1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative], the Department of Defence [W81XWH-10-1-0341], the Canadian Institutes of Health Research [CIHR] for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

ANECS recruitment was supported by project grants from the NHMRC [ID#339435], The Cancer Council Queensland [ID#4196615] and Cancer Council Tasmania [ID#403031 and ID#457636]. SEARCH recruitment was funded by a programme grant from Cancer Research UK [C490/A10124]. Stage 1 and stage 2 case genotyping was supported by the NHMRC [ID#552402, ID#1031333]. Control data were generated by the Wellcome Trust Case Control Consortium (WTCCC), and a full list of the investigators who contributed to the generation of the data is available from the WTCCC website. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02 - funding for this project was provided by the Wellcome Trust under award 085475. NSECG was supported by the EU FP7 CHIBCHA grant, Wellcome Trust Centre for Human Genetics Core Grant 090532/Z/09Z, and CORGI was funded by Cancer Research UK. Recruitment of the QIMR Berghofer controls was supported by the NHMRC. The University of Newcastle, the Gladys M Brawn Senior Research Fellowship scheme, The Vincent Fairfax Family Foundation, the Hunter Medical Research Institute and the Hunter Area Pathology Service all contributed towards the costs of establishing the Hunter Community Study.

The Bavarian Endometrial Cancer Study (BECS) was partly funded by the ELAN fund of the University of Erlangen. The Hannover-Jena Endometrial Cancer Study was partly supported by the Rudolf Bartling Foundation. The Leuven Endometrium Study (LES) was supported by the Verelst Foundation for endometrial cancer. The Mayo Endometrial Cancer Study (MECS) and Mayo controls (MAY) were supported by grants from the National Cancer Institute of United States Public Health Service [R01 CA122443, P30 CA15083, P50 CA136393, and GAME-ON the NCI Cancer Post-GWAS Initiative U19 CA148112], the Fred C and Katherine B Andersen Foundation, the Mayo Foundation, and the Ovarian Cancer Research Fund with support of the Smith family, in memory of Kathryn Sladek Smith. MoMaTEC received financial support from a Helse Vest Grant, the University of Bergen, Melzer Foundation, The Norwegian Cancer Society (Harald Andersens legat), The Research Council of Norway and Haukeland University Hospital. The Newcastle Endometrial Cancer Study (NECS) acknowledges contributions from the University of Newcastle, The NBN Children's Cancer Research Group, Ms Jennie Thomas and the Hunter Medical Research Institute. RENDOCAS was supported through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet [numbers: 20110222, 20110483, 20110141 and DF 07015], The Swedish Labor Market Insurance [number 100069] and The Swedish Cancer Society [number 11 0439]. The Cancer Hormone Replacement Epidemiology in Sweden Study (CAHRES, formerly called The Singapore and Swedish Breast/Endometrial Cancer Study; SASBAC) was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institutes of Health and the Susan G. Komen Breast Cancer Foundation.

The Breast Cancer Association Consortium (BCAC) is funded by Cancer Research UK [C1287/A10118, C1287/A12014]. The Ovarian Cancer Association Consortium (OCAC) is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith [PPD/RPCI.07], and the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.
Additional funding for individual control groups is detailed in the Supplementary Information

Table 1: Risk loci associated with EC at *P*< 5×10^{-8} in the meta-analysis.

Locus	SNP	Position	Nearby gene(s)	EA	OA	EAF	All histologies			Endometrioid histology		
							Allelic OR (95%CI)	*P*	I^2	Allelic OR (95%CI)	*P*	I^2
Novel GWAS loci												
13q22.1	rs11841589	73,814,891	*KLF5, KLF12*	G	T	0.74	1.15 (1.11-1.21)	4.83×10^{-11}	0.19	1.16 (1.10-1.21)	6.01×10^{-10}	0.00
6q22.31	rs13328298	126,016,580	*HEY2, NCOA7*	G	A	0.58	1.13 (1.09-1.18)	3.73×10^{-10}	0.00	1.15 (1.11-1.20)	1.02×10^{-11}	0.00
8q24.21	rs4733613	129,599,278	*MYC*	G	C	0.87	0.84 (0.80-0.89)	3.09×10^{-9}	0.00	0.84 (0.79-0.89)	7.70×10^{-9}	0.09
15q15.1	rs937213	40,322,124	*EIF2AK, BMF*	T	C	0.58	0.90 (0.86-0.93)	1.77×10^{-8}	0.36	0.90 (0.86-0.94)	2.22×10^{-7}	0.30
14q32.33	rs2498796	105,243,220	*AKT1, SIVA1*	G	A	0.70	0.89 (0.85-0.93)	3.55×10^{-8}	0.00	0.88 (0.85-0.92)	4.22×10^{-8}	0.00
Previously reported GWAS loci												
17q12	rs11263763	36,103,565	*HNF1B*	A	G	0.54	1.20 (1.15-1.25)	2.78×10^{-19}	0.37	1.20 (1.15-1.25)	6.51×10^{-17}	0.52
15q21	rs2414098	51,537,806	*CYP19A1*	C	T	0.62	1.17 (1.13-1.23)	4.51×10^{-13}	0.00	1.18 (1.13-1.23)	2.48×10^{-13}	0.00

Positions in build 37; EA, Effect allele; OA, Other allele; EAF, effect allele frequency; I^2, heterogeneity I^2 statistic[55]. For all novel loci, the lead SNP was either directly genotyped or imputed with an information score of more than 0.9. *HNF1B* and *CYP19A1* have been previously reported by Painter *et al.*[8] and Thompson *et al*[9].

Table 2: Conditional analysis of 8q24 locus showing two independent association signals.

SNP	Position	EA	OA	EAF	Pairwise r^2 with		All histology meta-analysis		Conditioning on rs4733613		Conditioning on rs17232730	
					rs4733613	rs17232730	Allelic OR (95%CI)	*P*	Allelic OR (95%CI)	*P*	Allelic OR (95%CI)	*P*
rs4733613	129,599,278	G	C	0.87	-	0.02	0.84 (0.79-0.89)	5.64×10^{-9}	-	-	0.86 (0.81-0.91)	$2.32 \times 10^{-}$
rs17232730	129,537,746	G	C	0.88	0.02	-	1.17 (1.10-1.24)	4.46×10^{-7}	1.14 (1.08-1.22)	1.29×10^{-5}	-	-
rs10088218*	129,543,949	G	A	0.87	0.02	0.43	1.14 (1.07-1.20)	1.65×10^{-5}	1.12 (1.05-1.18)	2.92×10^{-4}	1.01 (0.91-1.12)	0.818

Positions in build 37; EA, Effect allele; OA, Other allele; EAF, effect allele frequency.
*rs10088218 is associated with ovarian cancer (all subtypes), with the association being more significant for cancers of serous histology.
rs10088218-G is the risk allele for both EC and ovarian cancer.

References

1. Siegel, R., Ma, J., Zou, Z. & Jemal, A. Cancer statistics, 2014. *CA. Cancer J. Clin.* **64,** 9–29 (2014).
2. Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur. J. Cancer Oxf. Engl. 1990* **49,** 1374–1403 (2013).
3. Gruber, S. B. & Thompson, W. D. A population-based study of endometrial cancer and familial risk in younger women. Cancer and Steroid Hormone Study Group. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **5,** 411–417 (1996).
4. Win, A. K., Reece, J. C. & Ryan, S. Family history and risk of endometrial cancer: a systematic review and meta-analysis. *Obstet. Gynecol.* **125,** 89–98 (2015).
5. Barrow, E., Hill, J. & Evans, D. G. Cancer risk in Lynch Syndrome. *Fam. Cancer* **12,** 229–240 (2013).
6. Church, D. N. *et al.* DNA polymerase ε and δ exonuclease domain mutations in endometrial cancer. *Hum. Mol. Genet.* **22,** 2820–2828 (2013).
7. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **45,** 136–144 (2013).
8. Painter, J. N. *et al.* Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Hum. Mol. Genet.* (2014). doi:10.1093/hmg/ddu552
9. Thompson, D. J. *et al.* CYP19A1 fine-mapping and Mendelian randomisation: estradiol is causal for endometrial cancer. *Endocr. Relat. Cancer* (2015). doi:10.1530/ERC-15-0386
10. Spurdle, A. B. *et al.* Genome-wide association study identifies a common variant associated with risk of endometrial cancer. *Nat. Genet.* **43,** 451–454 (2011).
11. Tomlinson, I. *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39,** 984–988 (2007).

12. Tenesa, A. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.* **40,** 631–637 (2008).
13. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–678 (2007).
14. Pharoah, P. D. P. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.* **45,** 362–370, 370e1–2 (2013).
15. Sakoda, L. C., Jorgenson, E. & Witte, J. S. Turning of COGS moves forward findings for hormonally mediated cancers. *Nat. Genet.* **45,** 345–348 (2013).
16. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45,** 353–361 (2013).
17. Eeles, R. A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.* **45,** 385–391, 391e1–2 (2013).
18. De Vivo, I. *et al.* Genome-wide association study of endometrial cancer in E2C2. *Hum. Genet.* **133,** 211–224 (2014).
19. Setiawan, V. W. *et al.* Two estrogen-related variants in CYP19A1 and endometrial cancer risk: a pooled analysis in the Epidemiology of Endometrial Cancer Consortium. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **18,** 242–247 (2009).
20. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–585 (2013).
21. Cancer Genome Atlas Research Network *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497,** 67–73 (2013).
22. Katoh, M. & Katoh, M. Integrative genomic analyses on HES/HEY family: Notch-independent HES1, HES3 transcription in undifferentiated ES cells, and Notch-dependent HES1, HES5, HEY1, HEY2, HEYL transcription in fetal tissues, adult tissues, or cancer. *Int. J. Oncol.* **31,** 461–466 (2007).

23. Shao, W., Halachmi, S. & Brown, M. ERAP140, a conserved tissue-specific nuclear receptor coactivator. *Mol. Cell. Biol.* **22,** 3358–3372 (2002).
24. Rothman, N. *et al.* A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat. Genet.* **42,** 978–984 (2010).
25. Kiemeney, L. A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* **40,** 1307–1312 (2008).
26. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447,** 1087–1093 (2007).
27. Whiffin, N. *et al.* Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum. Mol. Genet.* **23,** 4729–4737 (2014).
28. Goode, E. L. *et al.* A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat. Genet.* **42,** 874–879 (2010).
29. Eeles, R. A. *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat. Genet.* **41,** 1116–1121 (2009).
30. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet.* **41,** 1122–1126 (2009).
31. Huppi, K., Pitt, J. J., Wahlberg, B. M. & Caplen, N. J. The 8q24 gene desert: an oasis of non-coding transcriptional activity. *Front. Genet.* **3,** 69 (2012).
32. Carvajal-Carmona, L. G. *et al.* Candidate locus analysis of the TERT-CLPTM1L cancer risk region on chromosome 5p15 identifies multiple independent variants associated with endometrial cancer risk. *Hum. Genet.* (2014). doi:10.1007/s00439-014-1515-4
33. Berlanga, J. J., Santoyo, J. & De Haro, C. Characterization of a mammalian homolog of the GCN2 eukaryotic initiation factor 2alpha kinase. *Eur. J. Biochem. FEBS* **265,** 754–762 (1999).
34. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).
35. Cantley, L. C. The phosphoinositide 3-kinase pathway. *Science* **296,** 1655–1657 (2002).

36. Slomovitz, B. M. & Coleman, R. L. The PI3K/AKT/mTOR pathway as a therapeutic target in endometrial cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18,** 5856–5864 (2012).
37. Salvesen, H. B. *et al.* Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4834–4839 (2009).
38. Shoji, K. *et al.* The oncogenic mutation in the pleckstrin homology domain of AKT1 in endometrial carcinomas. *Br. J. Cancer* **101,** 145–148 (2009).
39. Cohen, Y. *et al.* AKT1 pleckstrin homology domain E17K activating mutation in endometrial carcinoma. *Gynecol. Oncol.* **116,** 88–91 (2010).
40. Du, W. *et al.* Suppression of p53 activity by Siva1. *Cell Death Differ.* **16,** 1493–1504 (2009).
41. Wang, X. *et al.* Siva1 inhibits p53 function by acting as an ARF E3 ubiquitin ligase. *Nat. Commun.* **4,** 1551 (2013).
42. Li, N. *et al.* Siva1 suppresses epithelial-mesenchymal transition and metastasis of tumor cells by inhibiting stathmin and stabilizing microtubules. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 12851–12856 (2011).
43. Dun, B. *et al.* Mycophenolic acid inhibits migration and invasion of gastric cancer cells via multiple molecular pathways. *PloS One* **8,** e81702 (2013).
44. Shi, H., Zhang, Z., Wang, X., Liu, S. & Teng, C. T. Isolation and characterization of a gene encoding human Kruppel-like factor 5 (IKLF): binding to the CAAT/GT box of the mouse lactoferrin gene promoter. *Nucleic Acids Res.* **27,** 4807–4815 (1999).
45. Simmen, R. C. M. *et al.* The emerging role of Krüppel-like factors in endocrine-responsive cancers of female reproductive tissues. *J. Endocrinol.* **204,** 223–231 (2010).
46. Mutter, G. L. *et al.* Global expression changes of constitutive and hormonally regulated genes during endometrial neoplastic transformation. *Gynecol. Oncol.* **83,** 177–185 (2001).

47. Davis, H. *et al.* FBXW7 mutations typically found in human cancers are distinct from null alleles and disrupt lung development. *J. Pathol.* **224,** 180–189 (2011).
48. Nandan, M. O. *et al.* Krüppel-like factor 5 mediates cellular transformation during oncogenic KRAS-induced intestinal tumorigenesis. *Gastroenterology* **134,** 120–130 (2008).
49. Forbes, S. A. *et al.* The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet. Editor. Board Jonathan Haines Al* **Chapter 10,** Unit 10.11 (2008).
50. ENCODE Project Consortium *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
51. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).
52. Wang, Y., Li, X. & Hu, H. H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics* **103,** 222–228 (2014).
53. O'Mara, T. A. *et al.* Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. *Endocr. Relat. Cancer* **22,** 851–861 (2015).
54. Cheng, T. H. *et al.* Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Sci. Rep.* **5,** 17369 (2015).
55. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21,** 1539–1558 (2002).

Figure legends

Figure 1: EC meta-analysis Manhattan plot

Manhattan plot of $-\log_{10}$-transformed P-values from meta-analysis of 22 autosomes. There are seven loci surpassing genome wide significance including two known loci: 15q21 (*CYP19A1*) and 17q12 (*HNF1B*) and five novel loci: 6q22 (*NCOA7, HEY2*), 8q24 (*MYC*), 13q22 (*KLF5*), 14q32 (*AKT1, SIVA1*), 15q15 (*EIF2AK4, BMF*).

Figure 2: Forest plots of novel EC risk loci

The odds ratio and 95% confidence intervals of each study of the meta-analysis are listed and shown in the adjacent plot. The I^2 heterogeneity scores (all <0.4) suggest that there is no marked difference in effects between studies. The SNPs represented are: a) rs11841589 (13q22), b) rs13328298 (6q22), c) rs4733613 (8q24), d) rs17232730 (8q24, pairwise r^2 0.02 with rs4733613), e) rs937213 (15q15) and f) rs2498796 (14q32).

Figure 3: Regional association plots for the five novel loci associated with EC.

The $-\log_{10}$ P-values from the meta-analysis and regional imputation for three GWAS and eight iCOGS groups are shown for SNPs at: a) 13q22.1, b) 6q22, c) & d) 8q24, e) 15q15 and f) 14q32.33. The SNP with the lowest P-value at each locus is labeled and marked as a purple diamond, and the dot color represents the LD with the top SNP. The blue line shows recombination rates in cM/Mb. All plotted SNPs are either genotyped or have an IMPUTE info score of more than 0.9 in all datasets. **Supplementary Figure 4** displays similar regional association plots with a larger number of SNPs using a less stringent info score cut-off.

Figure 4: **The 13q22.1 EC susceptibility locus**

a) *Diagram showing the 16kb region (position 73,804,930- 73,820,618) around rs11841589, rs9600103 and correlated SNPs rs7981863, rs7988505 and rs7989799 (black marks).*
FAIRE and ChIP assays with anti-H3K4Me2 and anti-H4Ac antibodies for three EC cell lines ARK-2 (rs9600103-TT), Ishikawa (rs9600103-AA) and AN3CA (rs9600103-AA) are shown, with the y-axis displaying enrichment normalized to non-crosslinked genomic DNA/sonicated input DNA, relative to the *Rhodopsin* promoter as a negative control using the ΔΔCt method. DNaseI hypersensitivity site (DHS) density signal in ENCODE EC Ishikawa cells (**Supplementary Note**) are shown, from experiments with cell lines treated with estrogen and tamoxifen. 100 vertebrates conservation is also displayed. Vertical dotted line represents the position of rs9600103.

b) *3C experiment for KLF5-expressing Ishikawa cells.* Relative interaction frequencies between an *Nco*I restriction fragment containing risk SNPs rs9600103 and rs11841589 (bait fragment) with NcoI fragments across the region were calculated using qPCR with normalization to the signal from a control BAC 3C library and a non-interacting chromosomal region, using the ΔΔCt method. The graph shows the frequencies plotted against the fragment position on chromosome 13. A significant interaction is seen with the fragment containing a *KLF5* transcriptional start site in Ishikawa cells.

c) *Luciferase reporter assay to analyze the activity of 3kb fragments containing either rs9600103 or rs11841589 using the pGL3 promoter vector in Ishikawa cells*. Green arrows represent the low-risk alleles, and red arrows the high-risk alleles. Error bars represent the standard error of the mean. Data were normalized by subtraction of background luminescence and normalized to pGL4 Renilla activity. Luciferase activity in the rs9600103-A risk allele was more than double than that of the rs9600103-T protective allele (P=0.018). Paired t-tests between the different fragments also showed that the rs9600103-A high-risk allele has significantly higher expression compared with both rs11841589 alleles (0.045, 0.039) (**Supplementary Table 8**). Schematic diagram displays position on chromosome 13 of the fragment sequences and the arrows represents the position of the two SNPs.